



the globus alliance
www.globus.org

Globus Day at OGF21

Jennifer M. Schopf

University Chicago and Argonne National Lab

Slides available at

<http://dev.globus.org/wiki/Outreach/OGF21>

Globus Day at OGF 21

- 9:00-10:30 Overview and Data Management 1
 - Introduction, Jennifer Schopf, ANL
 - GridFTP, John Bresnahan, University of Chicago
 - OGSA-DAI, Amy Krause, EPCC
- 11:00-12:30 Data Management 2 and Security
 - Data Catalyst (DDM), Steve Tuecke, UnivaUD
 - RLS and DRS, Ann Chervenak, USC ISI
 - Globus Security, Rachana Ananthakrishnan, ANL
 - GAARDS, Stephen Langella, Ohio State University
- 1:30-3:00: Scheduling and Higher-level Tools
 - GridWay, Eduardo Huedo & Tino Vazquez, U. Madrid
 - Gavia, Daniel Vanderster, University Victoria
 - UCLA Grid Portal, Prakashan, UCLA
 - Medicus, Ann Chervenak, USC ISI
- 3:15-4:45: Application Development and Open Questions
 - Introduce, Scott Oster & Stephen Langella, OSU
 - Remote Appl. Virtualization Infra. (RAVi), R. Madduri, ANL
 - GT4 distribution and Incubators, Jennifer Schopf, ANL
 - Closing panel for community questions



the globus alliance
www.globus.org

Introduction to Globus

Jennifer M. Schopf



What is a Grid?

- Resource sharing
 - Computers, storage, sensors, networks, ...
 - Sharing always conditional: issues of trust, policy, negotiation, payment, ...
- Coordinated problem solving
 - Beyond client-server: distributed data analysis, computation, collaboration, ...
- Dynamic, multi-institutional virtual orgs
 - Community overlays on classic org structures
 - Large or small, static or dynamic



Why Is this Hard or Different?

- Lack of central control
 - Where things run
 - When they run
- Shared resources
 - Contention, variability
- Communication and coordination
 - Different sites implies different sys admins, users, institutional goals, and often socio-political constraints



So Why Do It?

- Computations that need to be done with a time limit
- Data that can't fit on one site
- Data owned by multiple sites
- Applications that need to be run bigger, faster, more



Key Common Features

- The size and/or complexity of the problem
- Collaboration between people in several organizations
- Sharing computing resources, data, instruments



Why Grids Now? — The Changing Nature of Science

Collaborative

Project focused, globally distributed teams, spanning organizations within and beyond company boundaries

Distributed & Heterogeneous

Each team member/group brings own data, compute, & other resources into the project

Data & Computation Intensive

Access to computing and data resources must be coordinated across the collaboration

Dynamic Research

Science being addressed is changing as larger data sets can be analyzed and access to additional resources is made possible

Infrastructure must adapt to this new reality



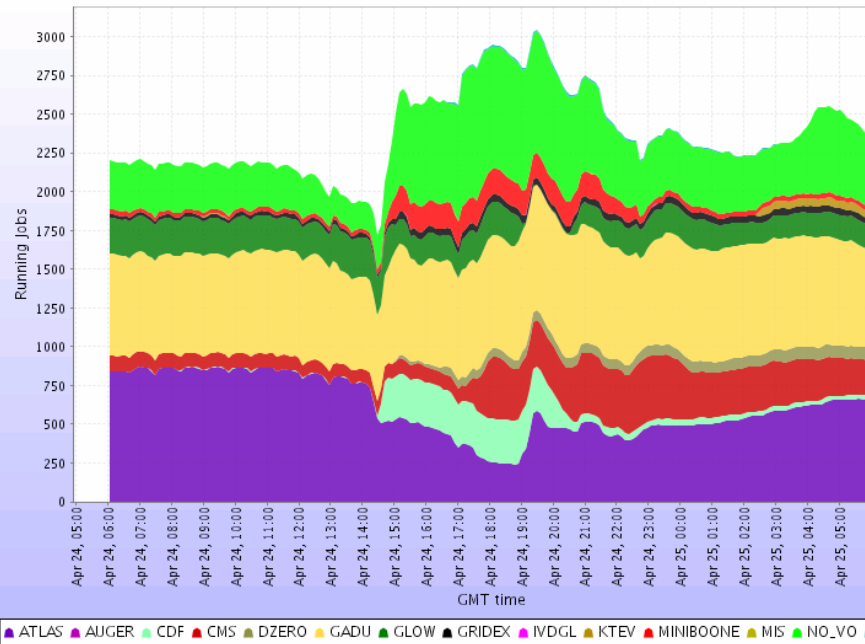
the globus alliance

www.globus.org

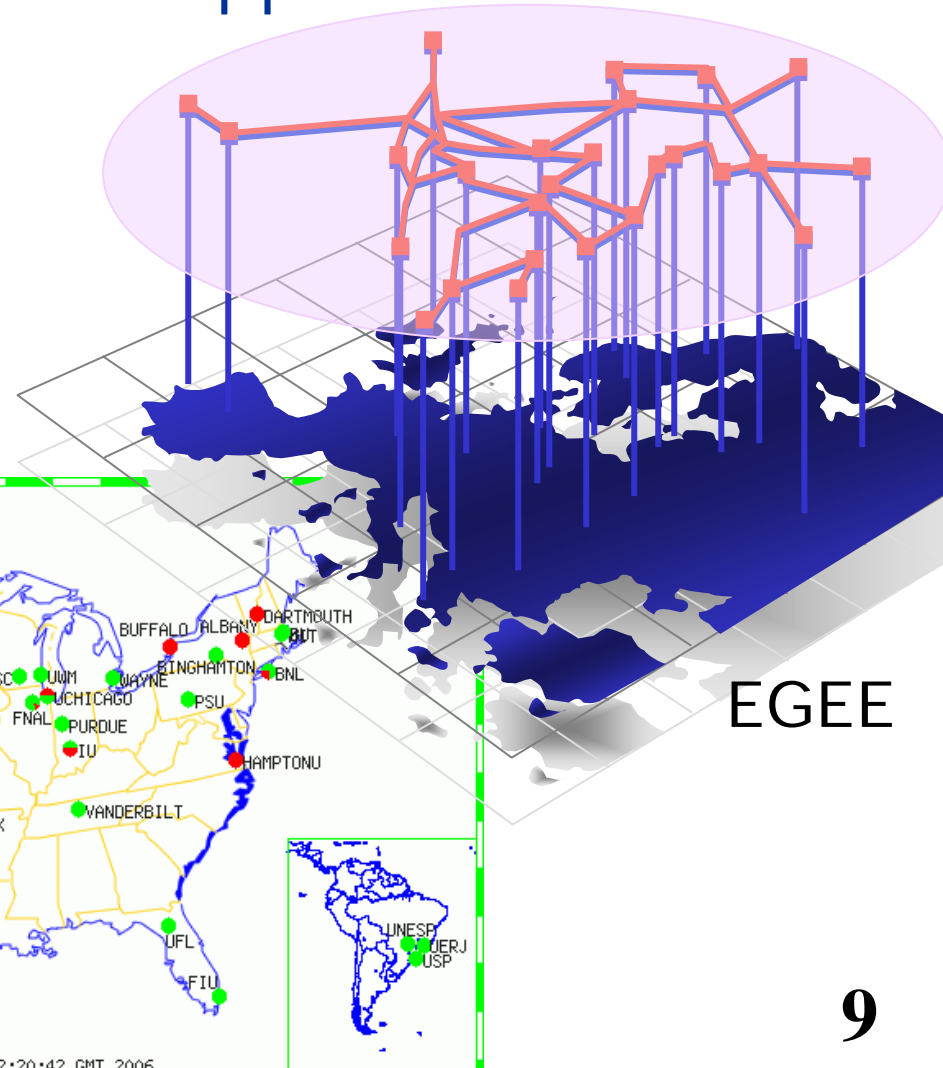
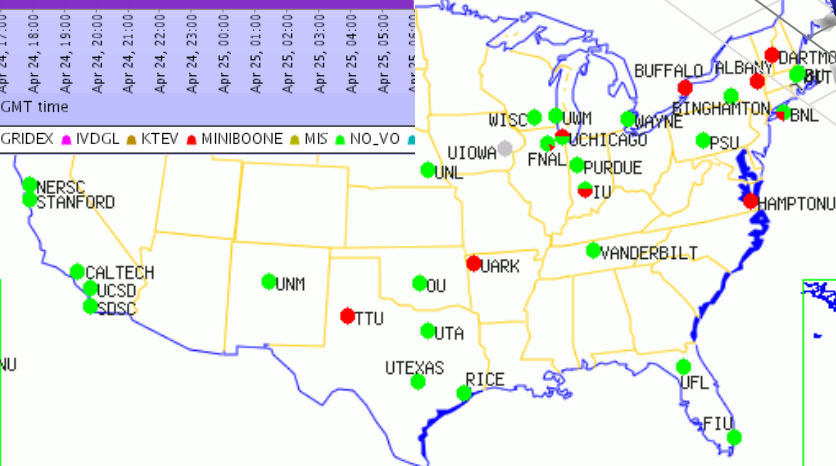
1st Generation Grids

Focus on aggregation of many resources for massively (data-)parallel applications

Running Jobs



Open Science Grid





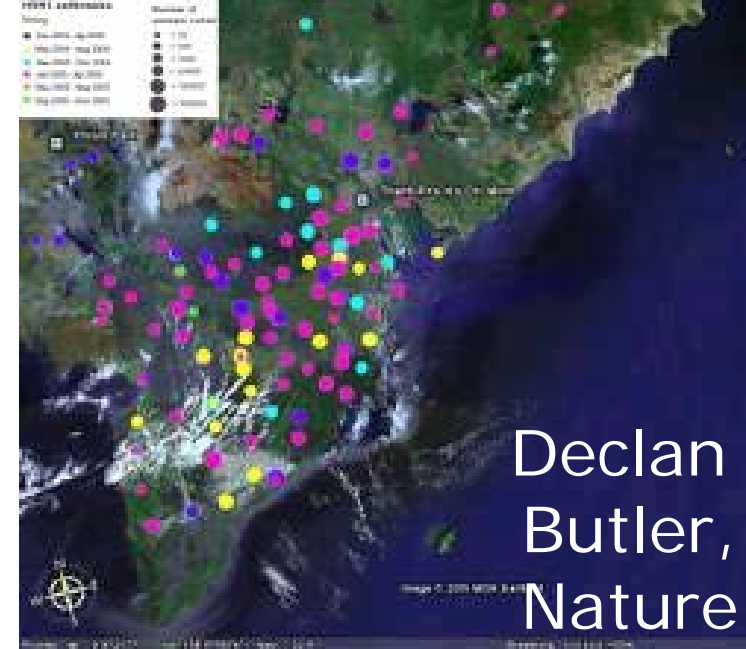
Second-Generation Grids

- Empower many more users by enabling on-demand access to **services**
- Science gateways (TeraGrid)
- Service oriented science
- Or, "Science 2.0"



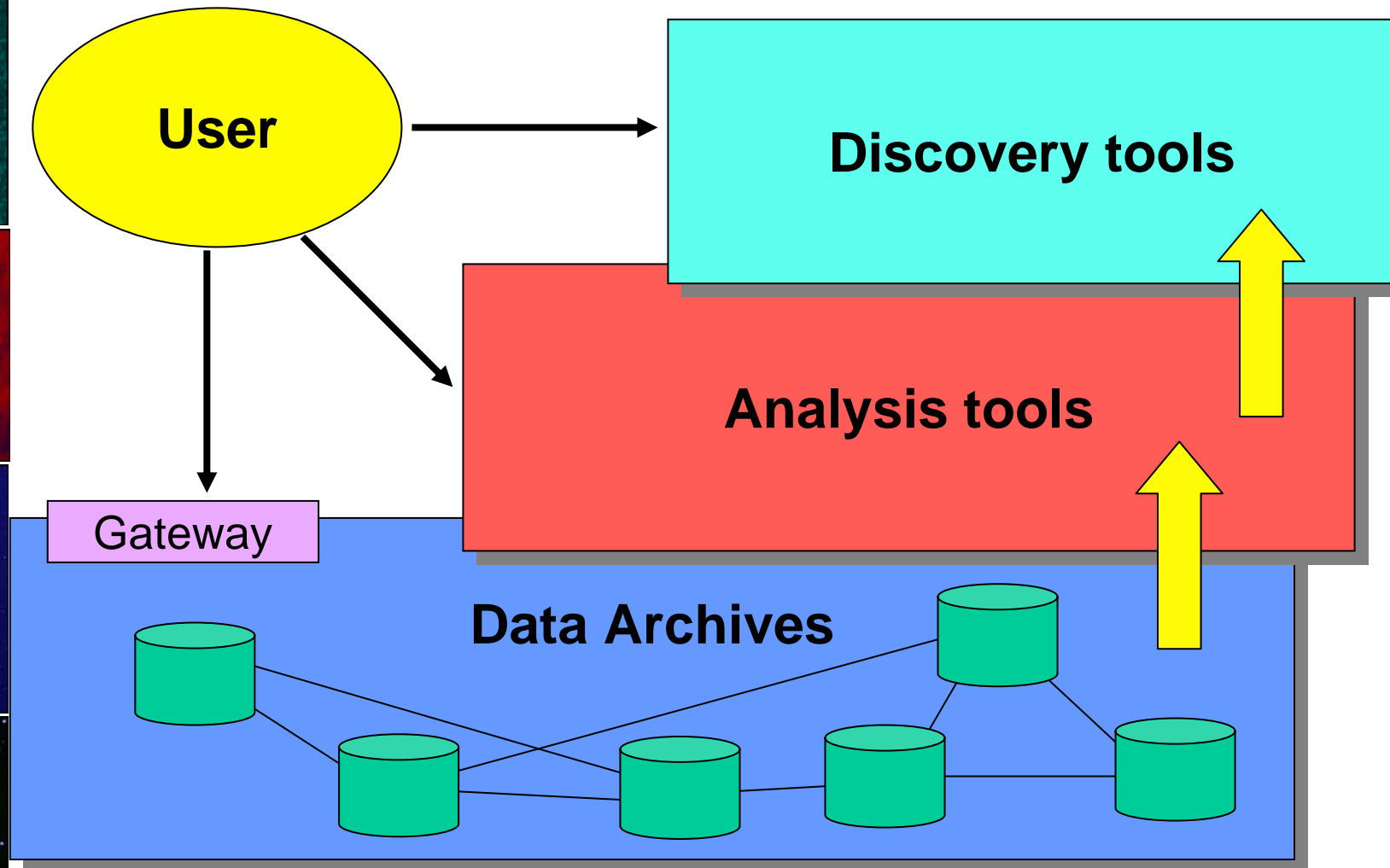
“Web 2.0”

- Software as services
 - Data- & computation-rich network services
- Services as platforms
 - Easy composition of services to create new capabilities (“mashups”)—that themselves may be made accessible as new services
- Participation beyond the creators
- Enabled by massive infrastructure buildout
 - Google projected to spend \$1.5B on computers, networks, and real estate in 2006
 - Many others are spending substantially





Science 2.0: E.g., Virtual Observatories





the globus alliance
www.globus.org

Service-Oriented Science

People **create** services (data or functions) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

→ I find "someone else" to **host** services,
so I don't have to become an expert in operating
services & computers!

→ I hope that this "someone else" can
manage security, reliability, scalability, ...

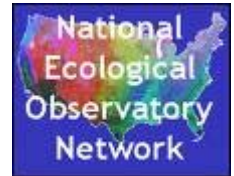
(Inter)national
Grids



Commercial Systems



Science Domain
Systems





Globus is...

- A collection of solutions to problems that come up frequently when building collaborative distributed applications
- Software for Grid infrastructure
 - Service enable new & existing resources
 - Uniform abstractions & mechanisms
- Tools to build applications that exploit Grid infrastructure
 - Registries, security, data management, ...
- Open source & open standards
 - Each empowers the other
- Enabler of a rich tool & service ecosystem



Globus is an Hour Glass

- Local sites have their own policies, installs – heterogeneity!

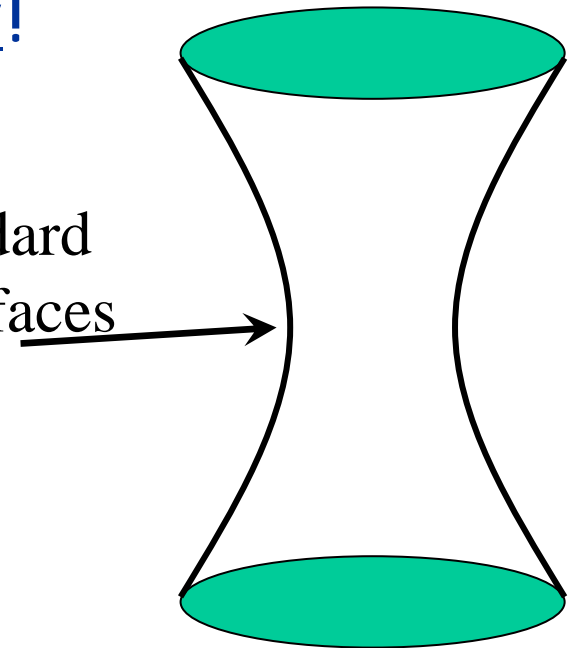
- Queuing systems, monitors, network protocols, etc

- Globus unifies – standards!

- Build on Web services
- Use WS-RF, WS-Notification to represent/access state
- Common management abstractions & interfaces

Higher-Level Services
and Users

Standard
Interfaces



Local heterogeneity



Globus is a Building Block

- Basic components for Grid functionality
 - Not turnkey solutions, but building blocks & tools for application developers & system integrators
- Highest-level services are often application specific, we let aps concentrate there
- Easier to reuse than to reinvent
 - Compatibility with other Grid systems comes for free
- We provide basic infrastructure to get you one step closer



Globus Philosophy

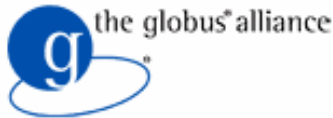
- Globus was first established as an open source project in 1996
- The Globus Toolkit is open source to:
 - Allow for inspection
 - > for consideration in standardization processes
 - Encourage adoption
 - > in pursuit of ubiquity and interoperability
 - Encourage contributions
 - > harness the expertise of the community
- The Globus Toolkit is distributed under the (BSD-style) Apache License version 2



- Governance model based on Apache Jakarta
 - Consensus based decision making
- Globus software is organized as several dozen “Globus Projects”
 - Each project has its own “Committers” responsible for their products
 - Cross-project coordination through shared interactions and committers meetings
- A “Globus Management Committee”
 - Overall guidance and conflict resolution



the globus alliance



the globus alliance

<http://dev.globus.org>

FAQ
Site Map
Contact

Home

Globus Alliance

Globus Toolkit

Grid Software

Grid Solutions

dev.globus

 Log in

article

discussion

edit

history

- Welcome
- List of projects
- Guidelines
- Infrastructure
- How to contribute
- Project ideas
- Mailing lists
- Globus events
- Recent changes
- dev.globus FAQ

common runtime projects

- C Core Utilities
- C WS Core
- CoG jglobus
- Core WS Schema
- Java WS Core
- Python Core
- XIO

data projects

- Data Replication
- GridFTP
- OGSA-DAI
- Reliable File Transfer
- Replica Location

execution projects

- GRAM
- GridWay
- MPICH-G2

information projects

- MDS4

Welcome

Globus was first established as an open source software project in 1996. At that time, the Globus development team has expanded from a few individuals to a distributed, international community. In response to this growth, the Globus community (the "Globus Alliance") established in October 2005 a new source development *infrastructure* and meritocratic *governance model*, which together make the process by which a developer joins the Globus community both easier and more transparent.

The Globus governance model and infrastructure are based on those of Apache Jakarta. In brief, the governance model places control over each individual software component (*project*) in the hands of its most active and respected contributors (*committers*), with a Globus Management Committee (GMC) providing overall guidance and conflict resolution. The infrastructure comprises *repositories*, *email lists*, *Wikis*, and *bug trackers* configured to support per-project communication, access and management.

For more information, see:

- The [Globus Alliance Guidelines](#), which address various aspects of the Globus governance model and the Globus community.
- A description of the Globus Alliance [Infrastructure](#), and known upcoming downtimes
- A list of current Globus projects.
- Information about [Globus community events](#).
- The [conventions and guidelines](#) that apply to contributions.

Guidelines
(Apache
Jakarta)

Infrastructure
(CVS, email,
bugzilla, Wiki,
licenses)

Projects
Include

...



Globus Technology Areas

- Core runtime
 - Infrastructure for building new services
- Security
 - Apply uniform policy across distinct systems
- Execution management
 - Provision, deploy, & manage services
- Data management
 - Discover, transfer, & access large data
- Monitoring
 - Discover & monitor dynamic services



Globus Projects

Globus Toolkit

MPICH G2

OGSA-DAI

Incubation
Mgmt

Java
Runtime

C
Runtime

Python
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-
OpenSSH

GridWay

GRAM

Data
Rep

GridFTP

Reliable
File
Transfer

Replica
Location

MDS4

GT4 Docs

Common
Runtime

Security

Execution
Mgmt

Data Mgmt

Info
Services

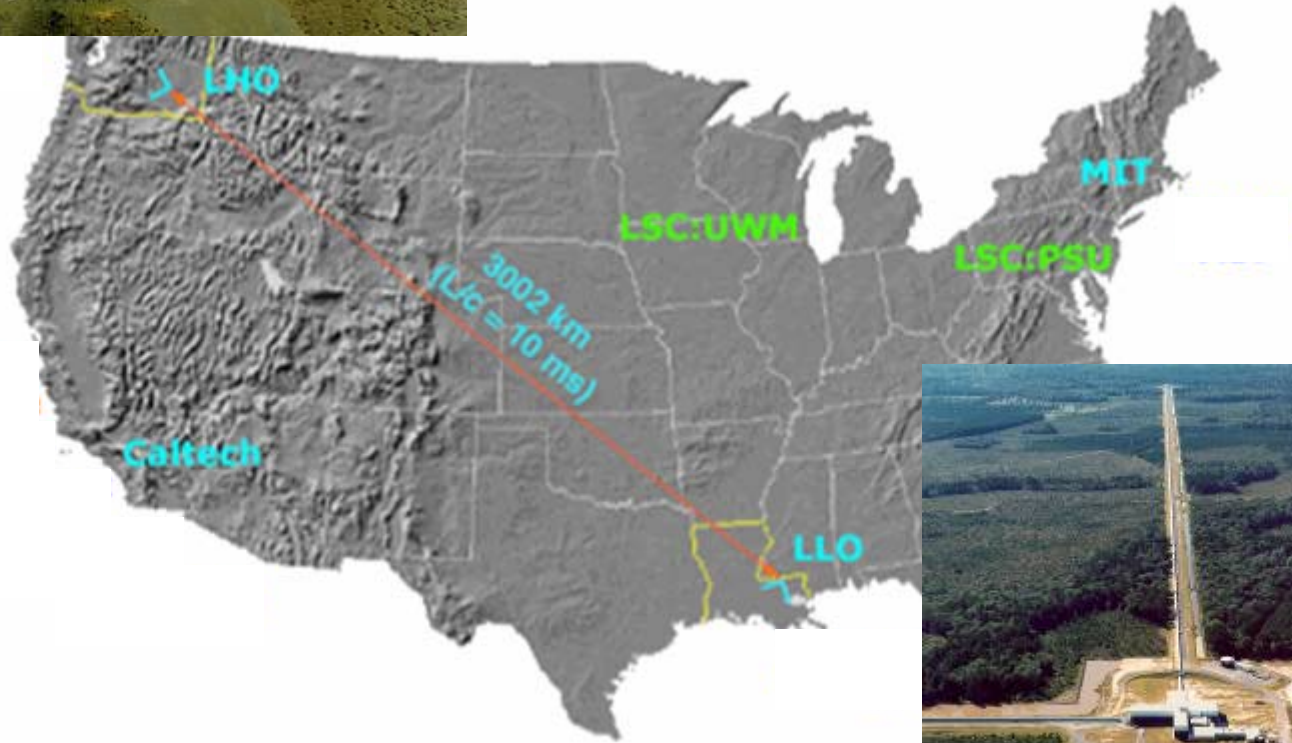
Other



the globus alliance
www.globus.org

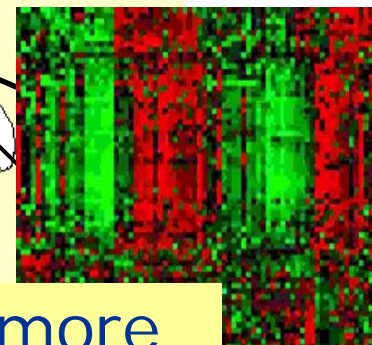
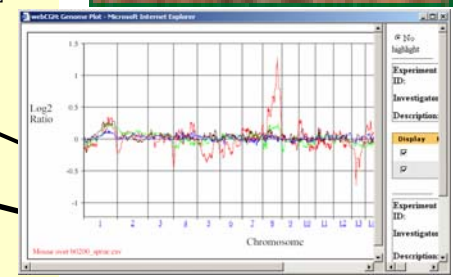
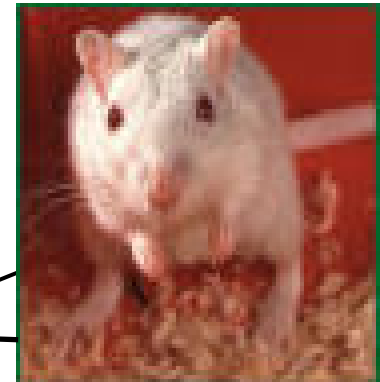
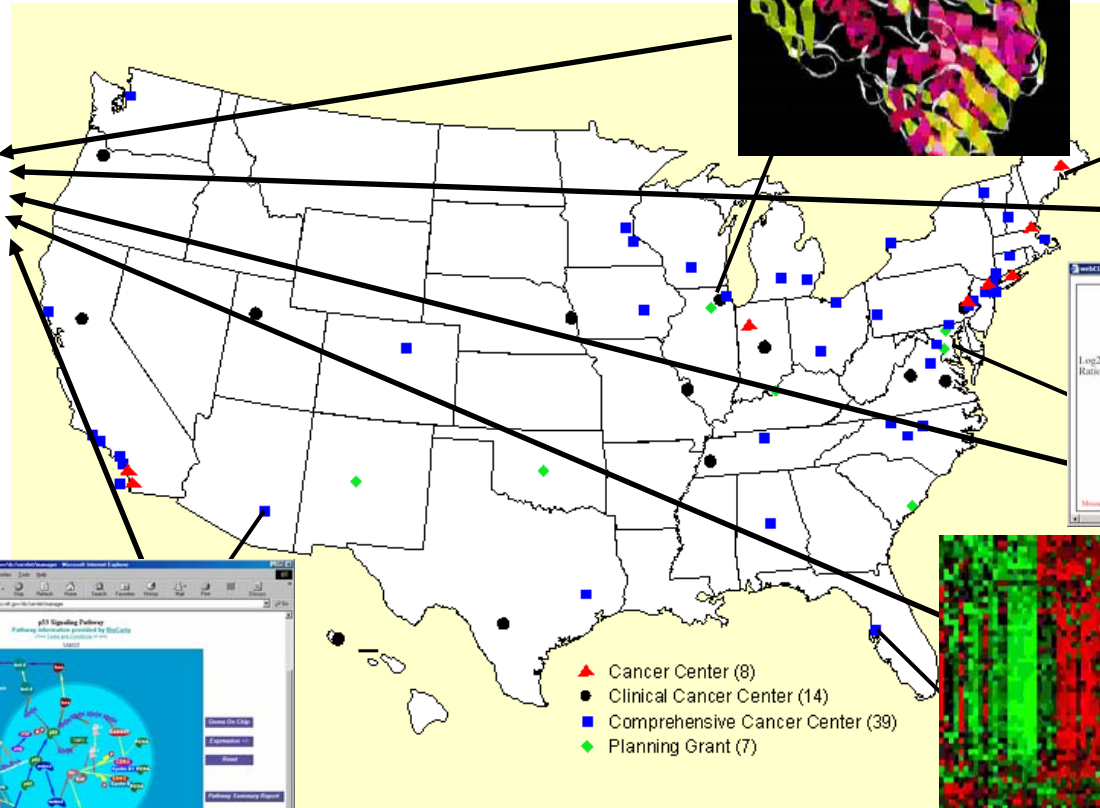
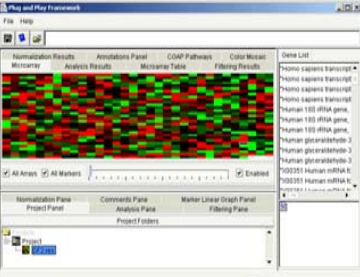
What do people do with Globus?

LIGO Gravitational Wave Observatory

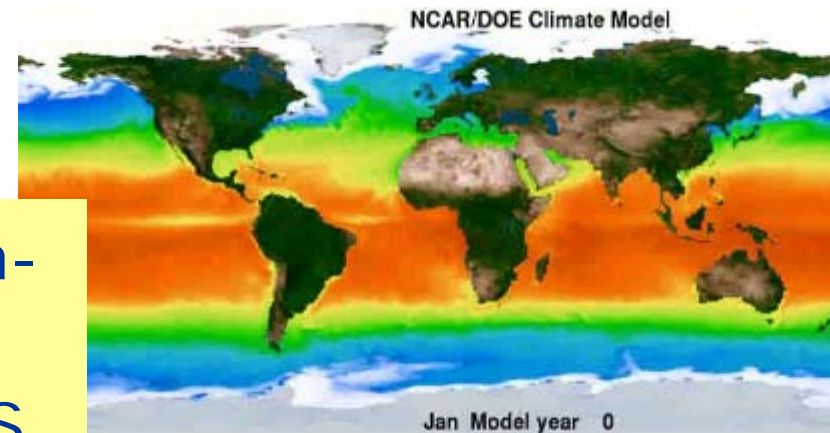
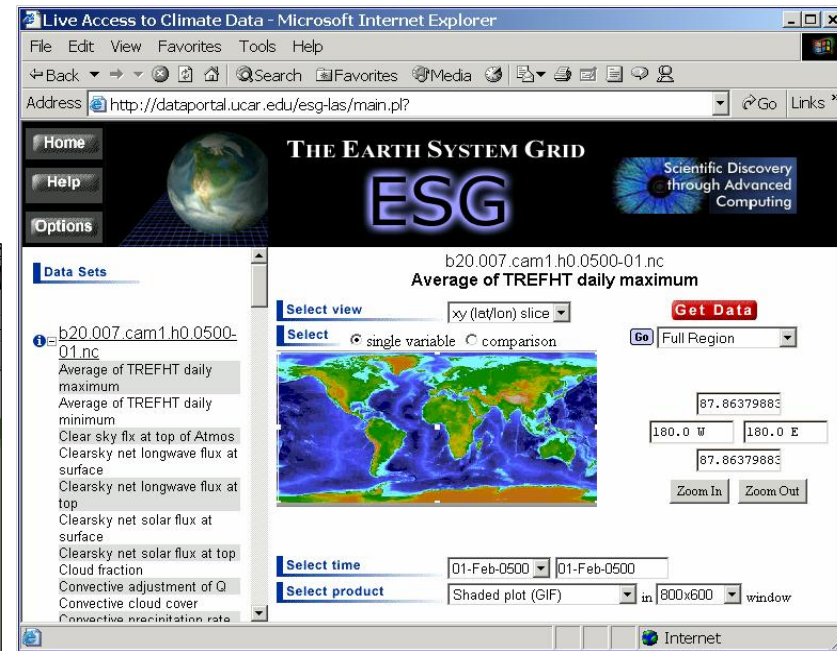
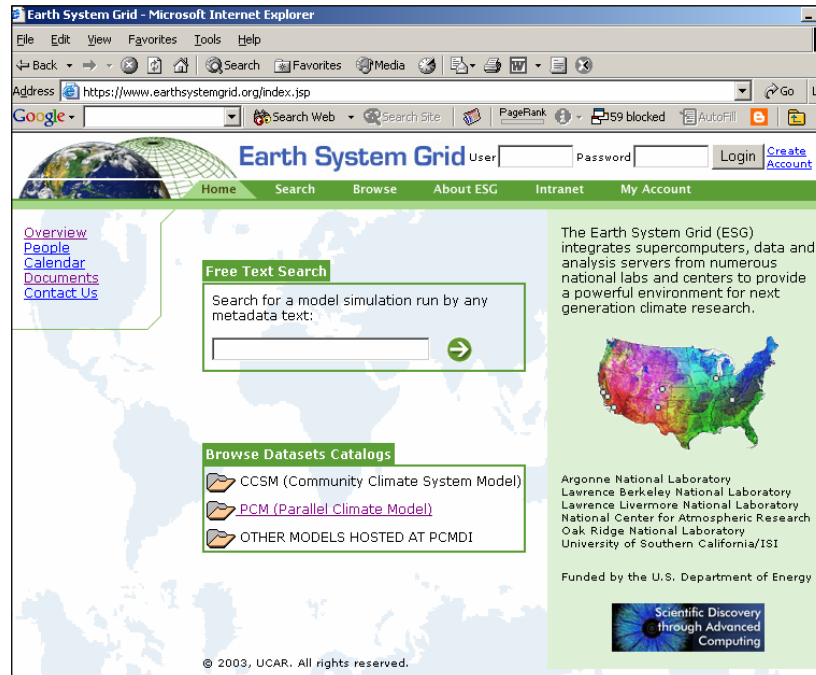


Replicate 1 TB/day of data to
10+ international sites
Uses GridFTP, RFT, RLS, DRS

CABig: Cancer Bioinformatics Grid



Integrate software and services across more than 50 sites
Uses Introduce, RAVi, MDS, GSI, Core



Enable sharing & analysis of high-volume data from advanced earth system models: Uses RLS, MDS, GridFTP, GSI



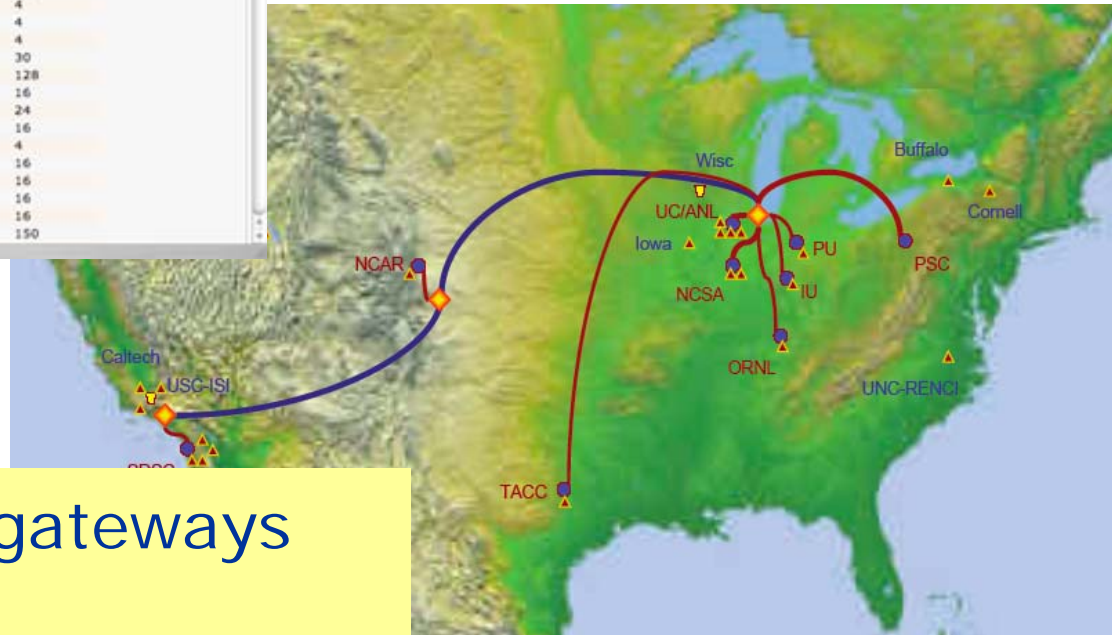
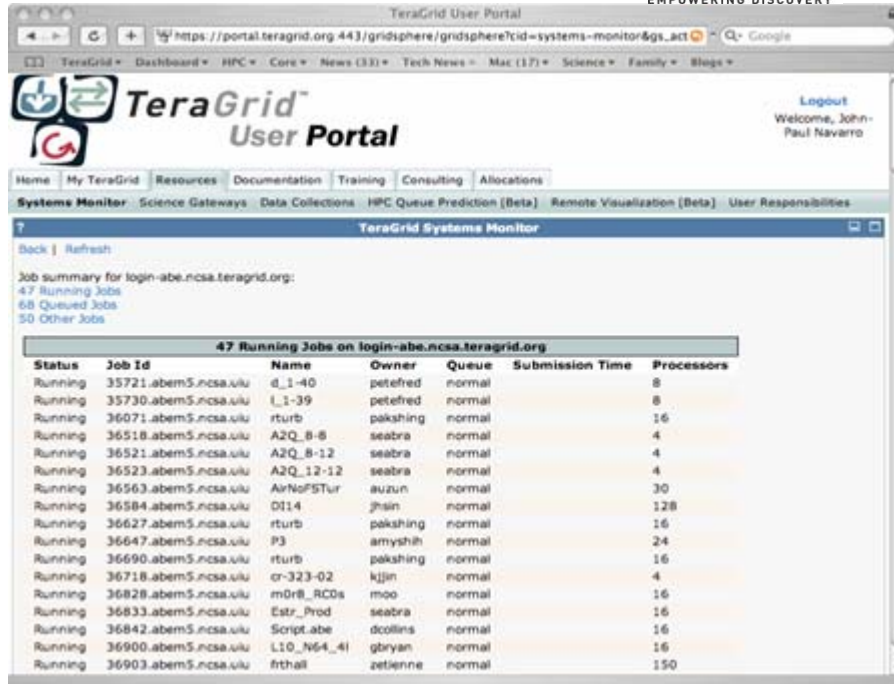
the globus alliance
www.globus.org



TeraGrid™
EMPOWERING DISCOVERY

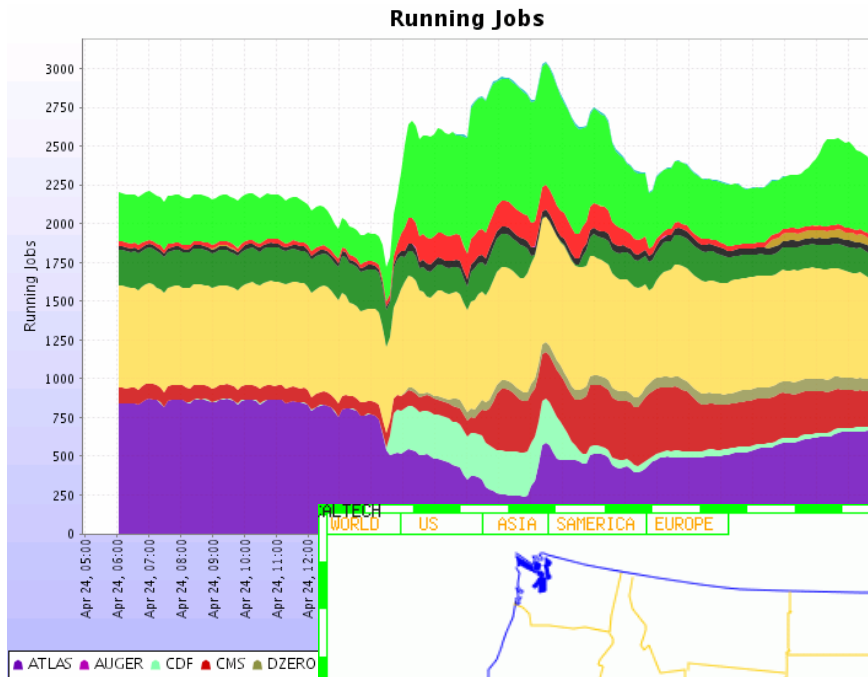
NSF's TeraGrid

- TeraGrid DEEP: Integrating NSF's most powerful computers (60+ TF)
- TeraGrid WIDE Science Gateways: Engaging Scientific Communities
- Base TeraGrid Cyberinfrastructure: Persistent, Reliable, National

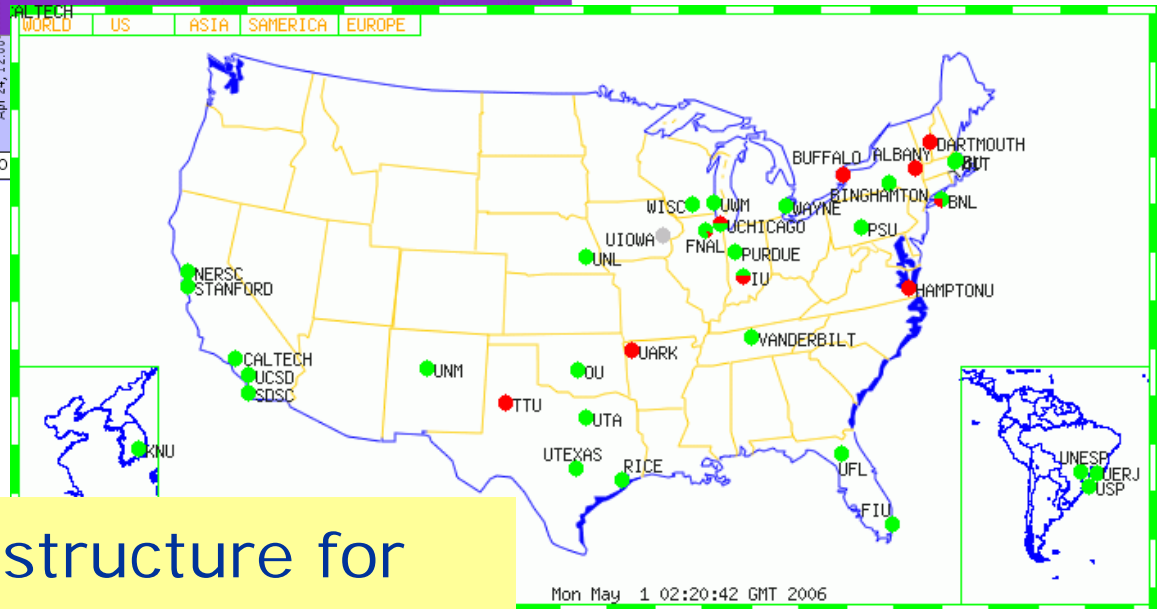


Enable user access and gateways
across NSF domains

Uses GRAM4, GridFTP, MDS, GSI, +



Open Science Grid



Enable common infrastructure for
DOE applications

Uses GRAM2, GridFTP, GSI, +



Globus Projects

MPICH G2

OGSA-DAI

Incubation
Mgmt

Java
Runtime

C
Runtime

Python
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-
OpenSSH

GridWay

GRAM

Data
Rep

GridFTP

Reliable
File
Transfer

Globus Toolkit

Replica
Location

MDS4

GT4 Docs

Incubator Projects

Common
Runtime

Security

Execution
Mgmt

Data Mgmt

Info
Services

Other



GT4 is a Distribution Project

- Collection of dev.globus software packaged and documented
- Components included must meet additional requirements
 - Coding standard
 - Testing coverage
 - Documentation coverage
 - > User, admin, developer
 - Response time for bugs and releases
 - Cannot change interfaces within a major release



Our Goals for GT4

- Usability, reliability, scalability, ...
 - Web service components have quality equal or superior to pre-WS components
 - Documentation at acceptable quality level
- Consistency with latest standards (WS-*, WSRF, WS-N, etc.) and Apache platform
 - WS-I Basic Profile compliant
 - WS-I Basic Security Profile compliant
- New components, platforms, languages
 - And links to larger Globus ecosystem



Versioning and Support

- Versioning
 - Evens are production (4.0.x, 4.2.x),
 - Odds are development (4.1.x)
- We support this version and the one previous
 - Currently stable version 4.0.5
 - We support 3.2.x and 4.0.x
 - We've also got the 4.1.2 dev release available (1 June '07)



Several “Next” Versions

- 4.0.6 – stable release
 - 100% same interfaces
 - Expected when there are “enough” bug fixes
- 4.1.3 – development release(s)
 - New functionality
 - Will include spec upgrade, so delayed until November-ish
- 4.2.0 - stable release
 - Tested, documented 4.1.x branch
 - Likely Q1 2008
 - Discussed on gt-dev@globus.org
- 5.0 – substantial code base change
 - With any luck, not for years :)



Globus Projects

MPICH G2

OGSA-DAI

Incubation
Mgmt

Java
Runtime

C
Runtime

Python
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-
OpenSSH

GridWay

GRAM

Data
Rep

GridFTP

Reliable
File
Transfer

Globus Toolkit

Replica
Location

MDS4

GT4 Docs

Incubator Projects

GDTE

GridShib

OGRO

Introduce

PURSE

HOC-SA

Swift

MEDICUS

UGP

LRMA

GEMICA

Cog WF

Dyn Acct

WEEP

RAVI

Virt WkSp

Gavia JSC

Gavia MS

DDM

SGGC

MonMan

NetLogger

Metrics

ServMark

Common
Runtime

Security

Execution
Mgmt

Data Mgmt

Info
Services

Other



Globus Projects

MPICH G2

OGSA-DAI

Incubation
Mgmt

Java
Runtime

C
Runtime

Python
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-
OpenSSH

GridWay

GRAM

Data
Rep

GridFTP

Reliable
File
Transfer

Globus Toolkit

Replica
Location

MDS4

GT4 Docs

Incubator Projects

| | | | | | | | | |
|--------------------|----------|--------|--------|----------|-----------|-----------|----------|--------|
| Incubator Projects | | | Swift | GEMICA | RAVI | | | MonMan |
| | | | GAARDS | MEDICUS | Cog WF | Virt WkSp | | |
| GDTE | GridShib | OGRO | UGP | Dyn Acct | Gavia JSC | DDM | Metrics | |
| Introduce | PURSE | HOC-SA | LRMA | WEEP | Gavia MS | SGGC | ServMark | |

Common
Runtime

Security

Execution
Mgmt

Data Mgmt

Info
Services

Other



Globus Projects

MPICH G2

OGSA-DAI

Incubation
Mgmt

Java
Runtime

Delegation

MyProxy

C
Runtime

CAS

GSI-
OpenSSH

Python
Runtime

C Sec

GridWay

GRAM

Globus Toolkit

Data
Rep

Replica
Location

GridFTP

MDS4

Reliable
File
Transfer

GT4 Docs

Incubator Projects

Swift

GEMICA

RAVI

MonMan

GAARDS

MEDICUS

Cog WF

Virt WkSp

NetLogger

GDTE

GridShib

OGRO

UGP

Dyn Acct

Gavia JSC

DDM

Metrics

Introduce

PURSE

HOC-SA

LRMA

WEEP

Gavia MS

SGGC

ServMark

Common
Runtime

Security

Execution
Mgmt

Data Mgmt

Info
Services

Other

Incubator Process in dev.globus

- Entry point for new Globus projects
- Incubator Management Project (IMP)
 - Oversees incubator process from first contact to becoming a Globus project

http://dev.globus.org/wiki/Incubator/Incubator_Process



25 Current Active Incubator Projects

- CoG Workflow
- Distributed Data Management (DDM)
- Dynamic Accounts
- Grid Authentication and Authorization with Reliably Distributed Services (GAARDS)
- Gavia-Meta Scheduler
- Gavia- Job Submission Client
- Grid Development Tools for Eclipse (GDTE)
- Grid Execution Mgmt. for Legacy Code Apps. (GEMLCA)
- GridShib
- Higher Order Component Service Architecture (HOC-SA)
- Introduce
- Local Resource Manager Adaptors (LRMA)
- MEDICUS (Medical Imaging and Computing for Unified Information Sharing)
- Metrics
- MonMan
- NetLogger
- Open GRid OCSP (Online Certificate Status Protocol)
- Portal-based User Registration Service (PURSe)
- Remote App Virtualization Infrastr. (RAVI)
- ServMark
- SJTU GridFTP GUI Client (SGGC)
- Swift
- UCLA Grid Portal Software (UGP)
- Workflow Enactment Engine Project (WEEP)
- Virtual Workspaces



How Can You Contribute? Create a New Project

- Do you have a project you'd like to contribute?
- Does your software solve a problem you think the Globus community would be interested in?
- Contact incubator-committers@globus.org



Contribute to an Existing Project

- Contribute code, documentation, design ideas, and feature requests
- Joining the mailing lists
 - *-dev, *-user, *-announce for each project
 - See the project wiki page at dev.globus.org
- Chime in at any time
- Regular contributors can become committers, with a role in defining project directions

http://dev.globus.org/wiki/How_to_contribute



Globus Software During Today

Globus Projects

MPICH G2

OGSA-DAI

Incubation
Mgmt

Java
Runtime

C
Runtime

Python
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-
OpenSSH

GridWay

GRAM

Data
Rep

GridFTP

Reliable
File
Transfer

**Globus
Toolkit**

Replica
Location

MDS4

GT4 Docs

Incubator Projects

GDTE
Introduce

GridShip
PURSE

OGRO
HOC-SA

UGP
LRMA

Dyn Acct
WEEP

Gavia JSD
Gavia MS

DDM
SGGC

Metrics
ServMark

GAARDS

MEDICUS

Cog WF

Virt WkSp

Swift

GEMICA

RAVI

MonMan

NetLogger

**Common
Runtime**

Security

**Execution
Mgmt**

Data Mgmt

**Info
Services**

Other



For More Information

- Jennifer Schopf
 - jms@mcs.anl.gov
 - <http://www.mcs.anl.gov/~jms>
- Globus Alliance
 - <http://www.globus.org>
- Dev.globus
 - <http://dev.globus.org>
- Upcoming Events
 - <http://dev.globus.org/wiki/Outreach>

Globus Day at OGF 21

- 9:00-10:30 Overview and Data Management 1
 - Introduction, Jennifer Schopf, ANL
 - GridFTP, John Bresnahan, University of Chicago
 - OGSA-DAI, Amy Krause, EPCC
- 11:00-12:30 Data Management 2 and Security
 - Data Catalyst (DDM), Steve Tuecke, UnivaUD
 - RLS and DRS, Ann Chervenak, USC ISI
 - Globus Security, Rachana Ananthakrishnan, ANL
 - GAARDS, Stephen Langella, Ohio State University
- 1:30-3:00: Scheduling and Higher-level Tools
 - GridWay, Eduardo Huedo & Tino Vazquez, U. Madrid
 - Gavia, Daniel Vanderster, University Victoria
 - UCLA Grid Portal, Prakashan, UCLA
 - Medicus, Ann Chervenak, USC ISI
- 3:15-4:45: Application Development and Open Questions
 - Introduce, Scott Oster & Stephen Langella, OSU
 - Remote Appl. Virtualization Infra. (RAVi), R. Madduri, ANL
 - GT4 distribution and Incubators, Jennifer Schopf, ANL
 - Closing panel for community questions